Cyber Security, Big Data and Risk

Mark Seward, Sr. Director, Security and Compliance, Splunk

In-Depth Seminars – D24

ISACA®
Trust in, and value from, information systems
San Francisco Chapter

AGENDA

- Why are attacks successful?
- How does ‘big data’ help
- Changing our thinking
- The advanced threat ‘playbook’
- Thinking security – talking business risk
- Questions

Advanced threats are hard to detect

100%

Valid credentials were used

243

Median # of days before detection

40

Average # of systems accessed

63%

Of victims were notified by external entity

Source: Mandiant M-Trends Report 2012 and 2013

‘Attacker think’

Attackers don’t want to work too hard to get what they want.

“What’s the easiest way to target the right people who have access to the stuff I can sell?”

ITEM STOLEN	HOW THE ATTACKERS USE INFORMATION
 Network Infrastructure Documentation Including Schematics and Configuration Files	Understand firewall and other IDS configurations and where vulnerabilities that can be exploited exist.
 Organization Chart	Establish individuals to target in spear-phishing campaigns or to target for email and data theft.
 Systems Documentation	Identify where targeted systems existing within a victim network.
 VPN Configuration Files	Identify what VPN users have access to within a victim’s network and target VPN credential data to steal.

Why are attacks successful -- Silos

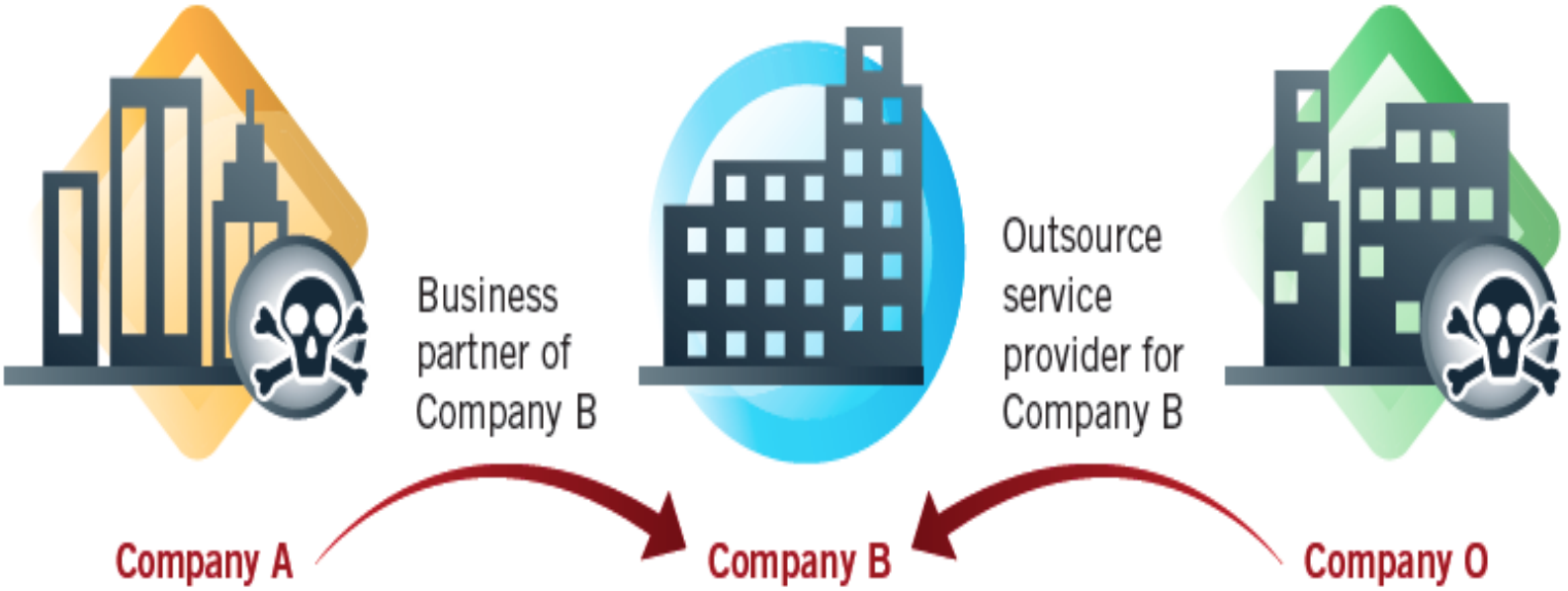
- Defenders are isolated focused on narrow defensive zones

- Opponents are organized, persistent and creative

Why are attacks successful – People

- They are the weak point in our cyber defense.
- Only takes one time to be right
- Employee activities are credentialed
- Bypassing the perimeter
- Must gain an understanding of what is normal and what is not.
- Need a real-time big data approach to security and statistical analysis of the data

Why are attacks successful – Your Partners

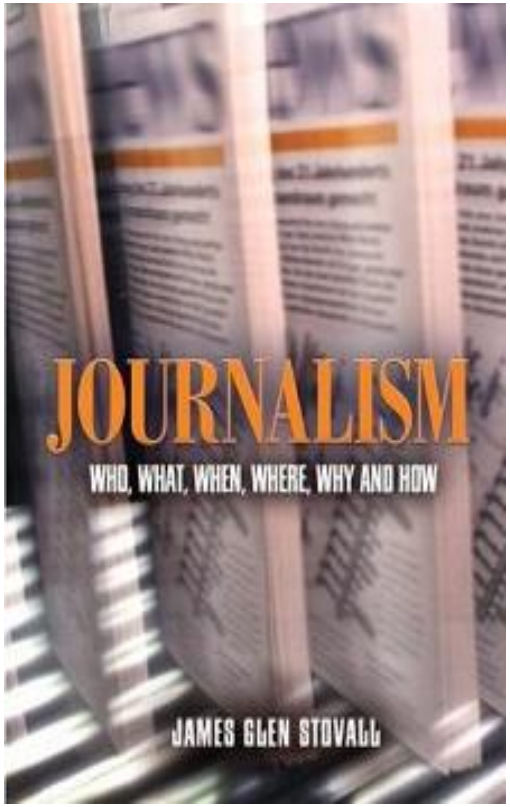


- Monitoring the partner and service provider access is about what's normal and what's not
- Understand your partner's cyber posture and policy

How much and what kinds of data do we need?

Telling your data security story

The 5 Ws of Journalism



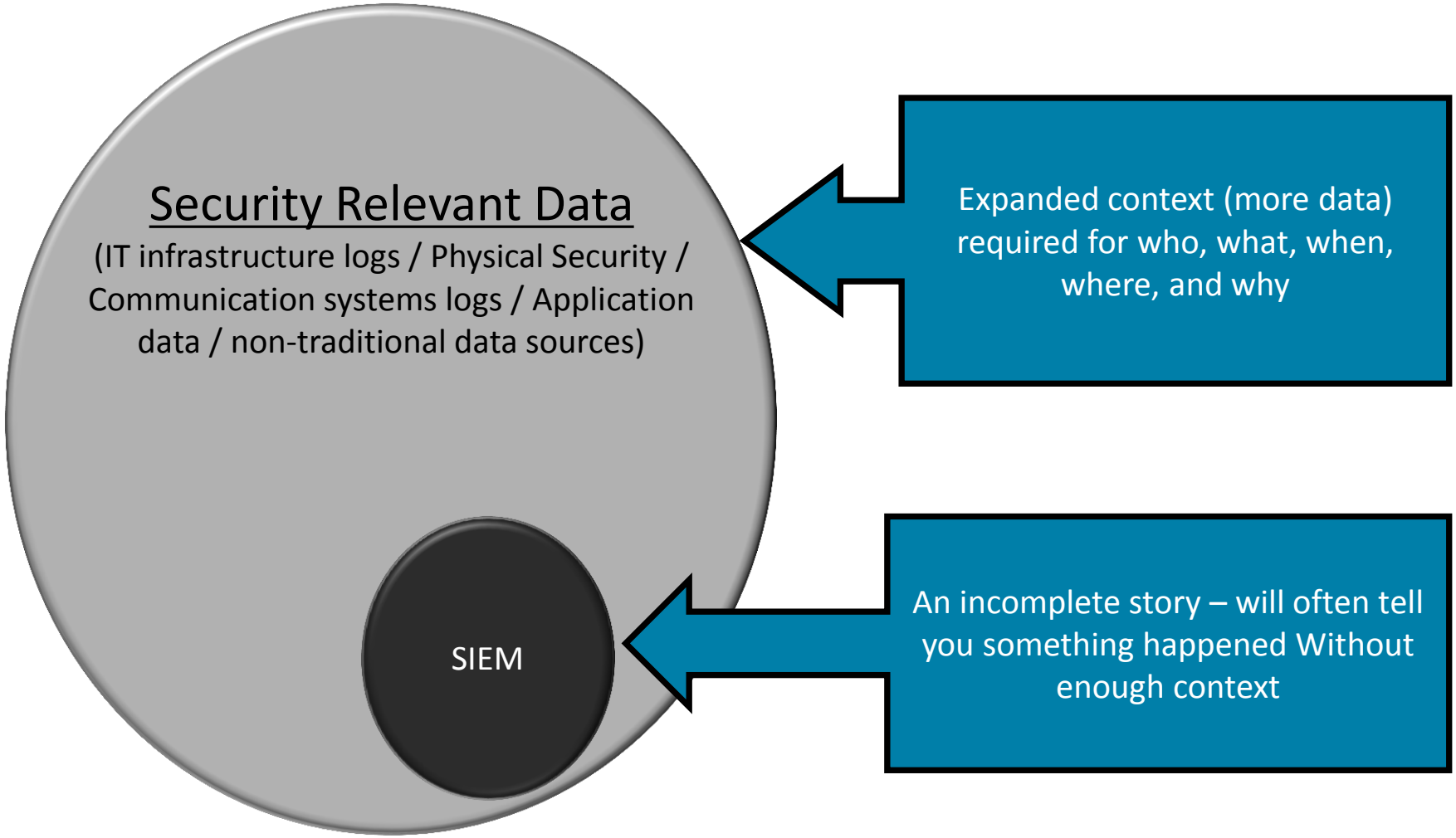
The 5 Ws of Information Security

Unstructured industrial control data: A risk blind spot

Security teams not focused on machine generated data

- 'Machines' deliver goods or services
- Machines monitor product quality
- Machine 'health' affects product/service quality
- Industrial Control Systems support JiT supply chains
- Environmental control data

An ever expanding universe of security data



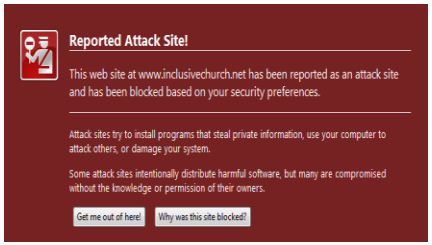
An ever expanding universe of security data

Email (How)

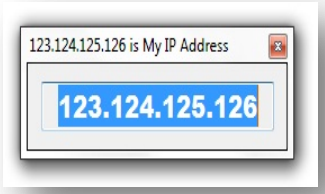
Time
(When)

Location
(Where)

Proxy data
Browsing
History
(What)



IP Address
DHCP / DNS
(Who)



Badge (Who)

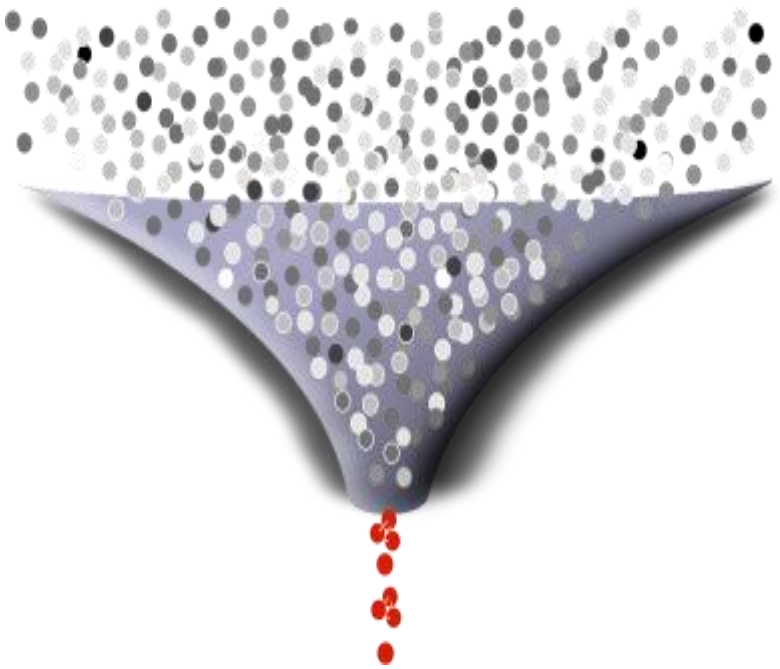
Supervisor (Why)

Date
(When)

The False Promise of SIEM and Data Reduction

Why are attacks successful – Data reduction

Typical SIEM Architecture



Data Reduction Model

- Have to know what you need for investigation before you need it
- Useful data can come from anywhere – not just what's supported by the vendor
- Lack of scalability restricts visibility
- Creates vendor dependency (people forget how to wade into their data)
- The 'cold case' problem

Security posture homogenized

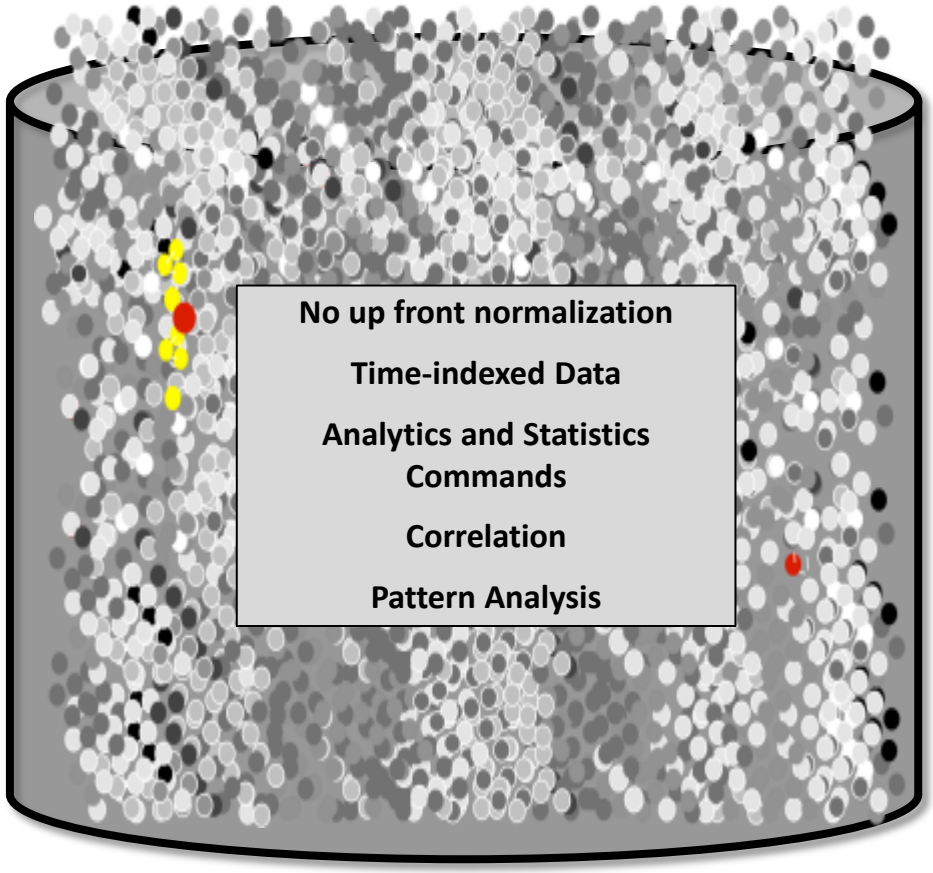
- Data reduction and normalization at collection time gives analysts a ‘Skim Milk’ view of security posture
- The ‘data fat’ can be relevant to an investigation
- All data is relevant for security

Moving to a data inclusion model

Specific behavior based pattern modeling for humans and machines

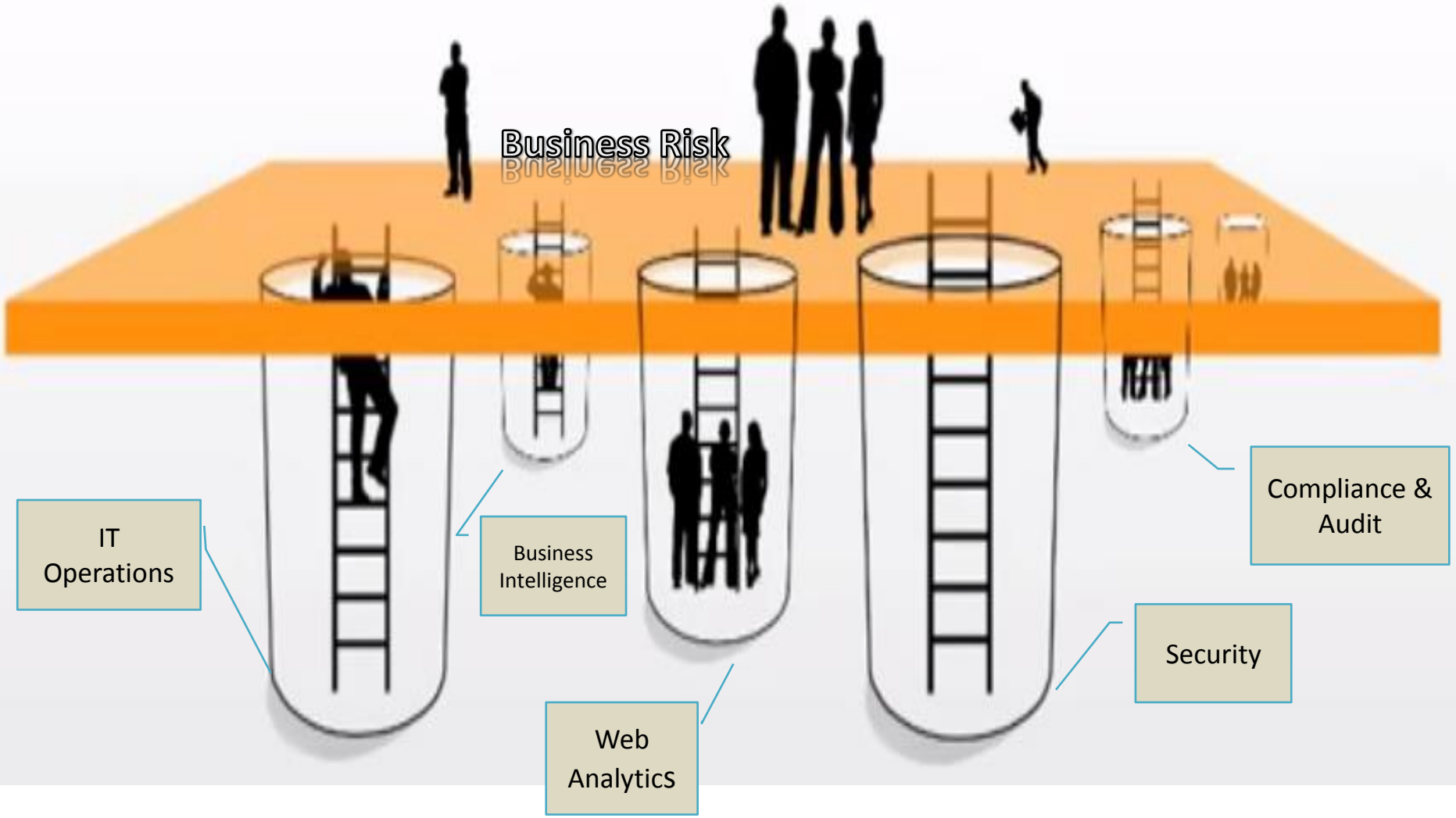
Based on combinations of:

- Location
- Role
- Data/Asset type
- Data/Asset criticality
- Time of day
- Action type
- Action length of time



Data Inclusion Model

Meeting at the big data layer



What's the playbook for advanced persistent attackers?

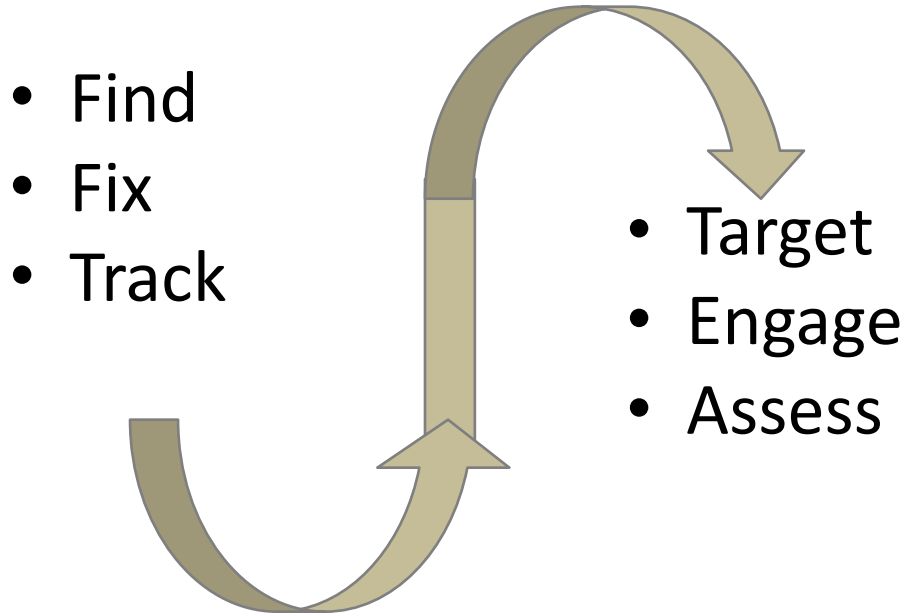
What is the Kill Chain?

- Represents the typical phases of an “advanced attack”
- What are the characteristics of an advanced threat or attack
  - Stealth
  - Stay resident as long as possible
  - Collection of ‘high value’ data
  - Can be nation state driven
  - Malware acts as a proxy for the malicious insider
  - Hacking the human – trust

The Kill-chain is a game film of typical attack activities –
a list of things that almost always happen but maybe not in order.

Kill-chain idea origin

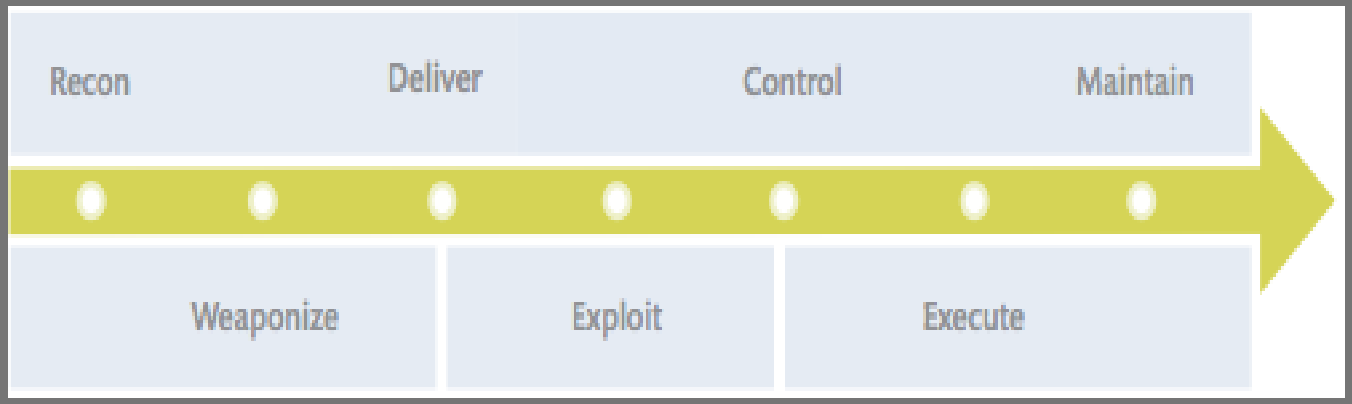
- In military parlance, a “Kill Chain” is a phase-based model to describe the stages of an attack, which also helps inform ways to prevent such attacks. These stages are referred to as:



The further towards the beginning of the Kill Chain an attack can be stopped, the better.

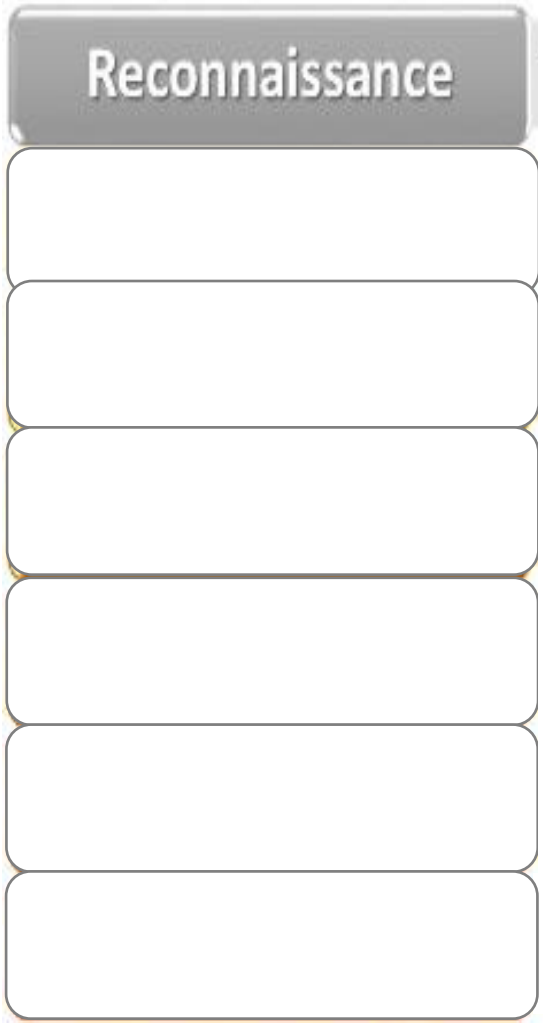
Kill-chain for cyber security as outlined by Miter

A successful strategy requires analysis of the “game film” called the advanced threat kill-chain



‘Kill-Chain’ activities defined

Monitoring the 'Kill-chain'



Web Analytics

- Get an understanding of clicks to the management or board member portion of your website from outside the country where your company is based.
- Google Analytics visitor flow report can help you understand where visitors come from how they troll and access the site.

Social Media

- Monitor out-going data, especially file sharing that may help an attacker with social engineering
- Monitor company sentiment to understand whether a "storm is gathering" that may result in an attack

Traffic Data

- originating from data center (know IP address spaces)

Monitoring the ‘Kill-chain’

Identify Threat Characteristics

- Identify the domain the email came from as a legitimate business
- Use analytics to understand if the email is seen for the first time from the sender.
- Monitor the types of attachments and perform packet level inspection to understand file attachment content (what is the attachment? Javascript, .exe, or does it contain a launch action)

Monitoring the ‘Kill-chain’

Malware Behavior Identification and Detection

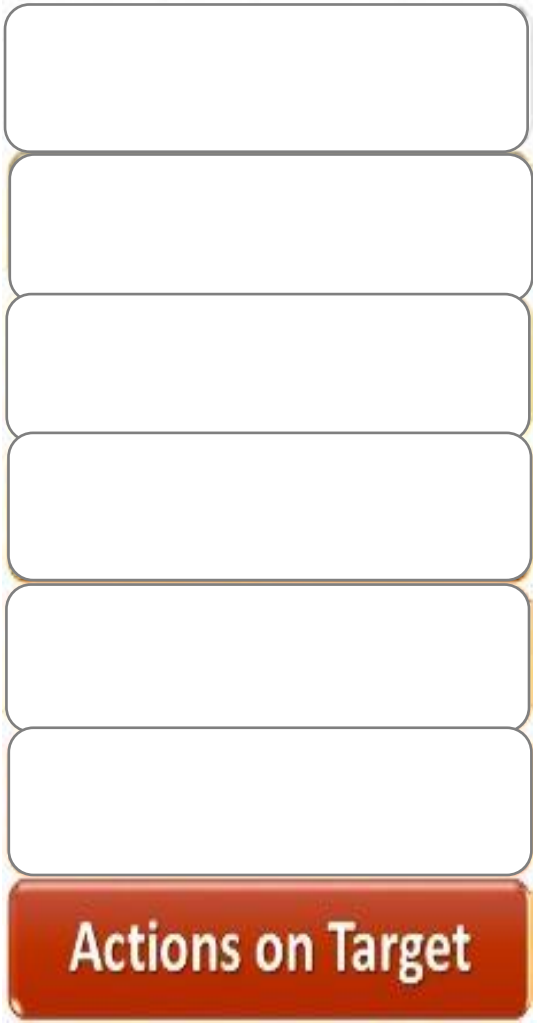
- Use Virus Total or GTRI’s Apiary to identify malware actions and characteristics
- Import Data from services into Splunk to monitor for infection characteristics not detected by AV engines
  - Collect malware Hash
  - Communication IPs, ports and protocols used
  - File or registry key changes
  - Domain the email came from as a legitimate business
  - Network connection(s)
  - DLL changes
- Correlate this data with host data collected
  - Are changes made outside of change windows
- Monitor for unusual rare traffic between hosts for lateral movement
- Monitor changes to hosts processes

Monitoring the ‘Kill-chain’

Malware Communication Analytics

- Monitor URL / and user agent strings for embedded command and control
  - Lengths above particular standard deviation
- Monitor web traffic to known bad IPs and domains
- Monitor web traffic to domains registered in the last 24-72 hours
- Monitor web traffic w/o referrer
- Use Virus Total or GTRI’s Apiary to identify malware actions and characteristics
- Outbound encrypted traffic (from DMZ, webservers, DBs, other hosts that should not be initiating connections
- Identify self-signed certificates
- Falsified HTTP headers
- Beaconing hosts
- Non-standard encryption over allowed paths
- Use of Remote windows shell or remote desktop

Monitoring the ‘Kill-chain’



- DDoS from the inside
- CPU cycles eaten up
- Performance degradation
- Land and expand (what hosts are exhibiting same issues)
- Webserver content replaced
- Log files missing/erased
- New executable on host
- Host AV not updating
- Elevated privileges
- Movement of encrypted .rar or .zip files
- Use of sftp or ftp to a controlled host
- Use of pwdump tool

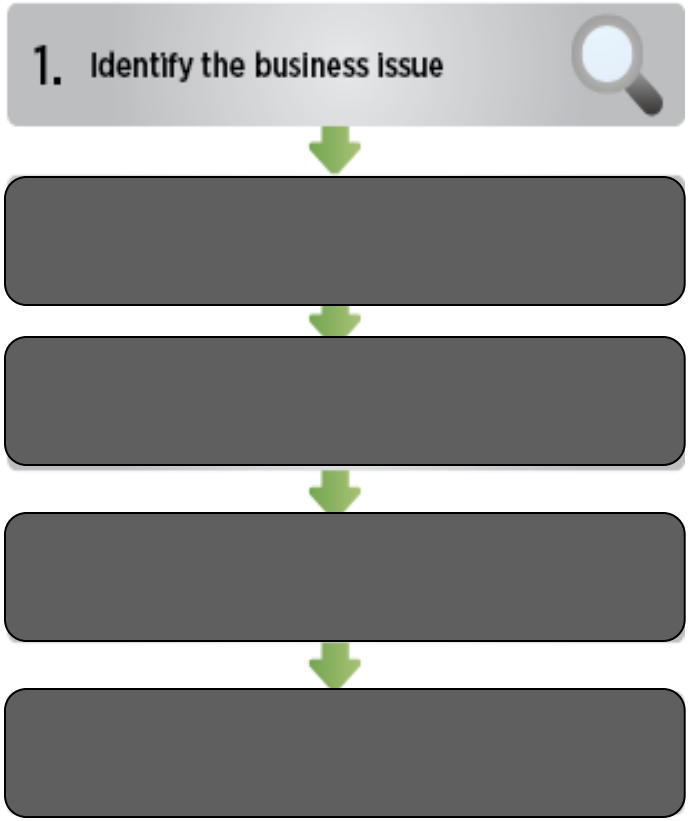
A tall order for the average security team?

- Take small measures/steps
- Pick one phase and focus – then pick the next one
  - Stopping the attacker at any one phase is good
- The earlier in the chain you are able to focus – the better
- Know your environment – you can bet the attacker will try to know it
- What information does your web presence tell an attacker?

Don’t let vendors tell you what questions their solution can answer.

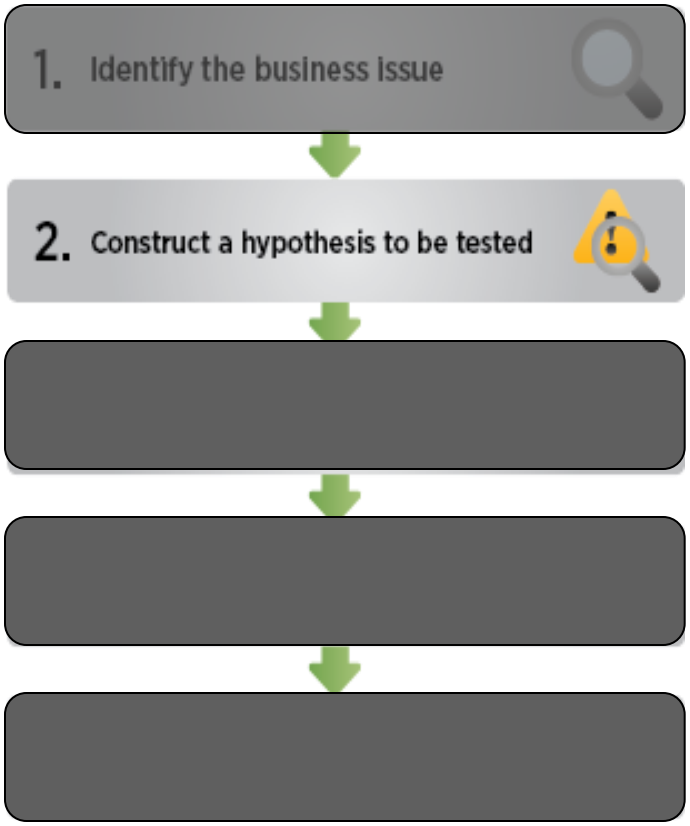
Ask the questions your business cares about.

A Process for Using Big Data for Security: Identify the Business Issue



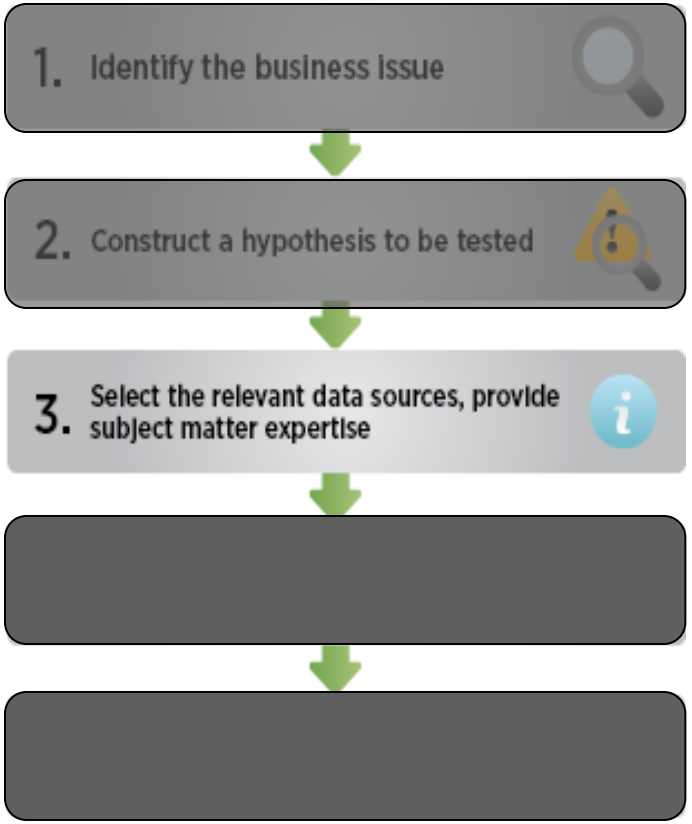
- What does the business care about?
- What could cause loss of service or financial harm?
- Performance Degradation
- Unplanned outages (security related)
- Intellectual property access
- Data theft

A Process for Using Big Data for Security: Construct a Hypothesis



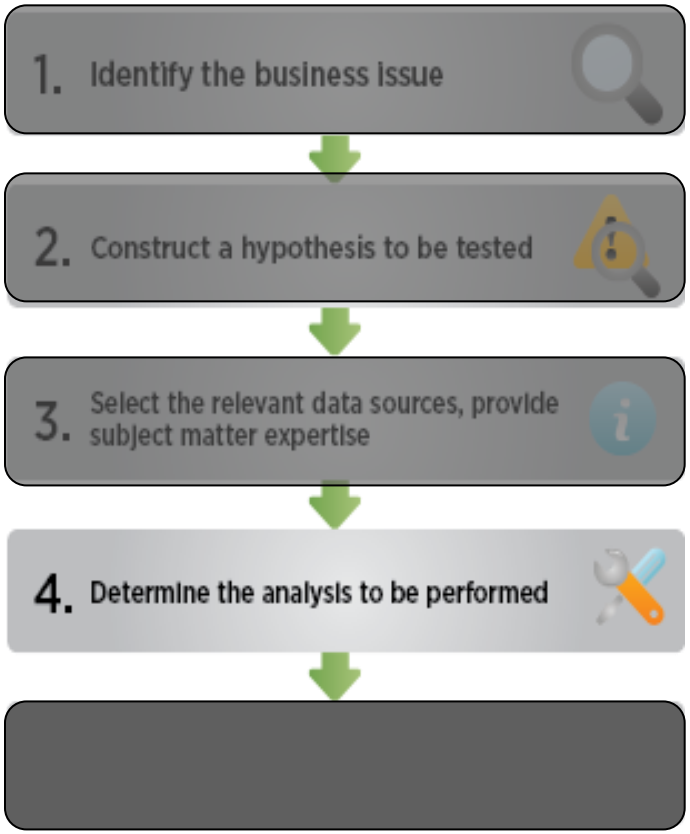
- How could someone gain access to data that should be kept private?
- What could cause a mass system outage does the business care about?
- What could cause performance degradation resulting in an increase in customers dissatisfaction?

A Process for Using Big Data for Security: It's about the Data



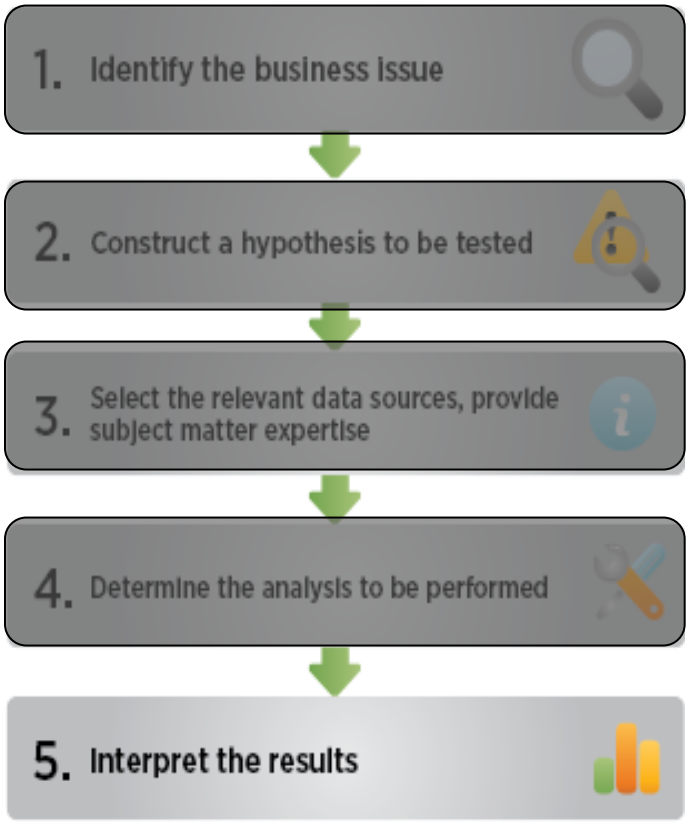
- Where might our problem be in evidence?
- For data theft start with unauthorized access issues…
- Facility access data, VPN, AD, Wireless, Applications, others…
- Beg, Borrow, SME from system owners

A Process for Using Big Data for Security: Data Analysis



- For data theft start with what's normal and what's not (create a statistical model)
- How do we 'normally' behave?
- What patterns would we see to identify outliers?
- Patterns based on ToD, Length of time, who, organizational role, IP geo-lookups, the order in which things happen, how often a thing normally happens, etc.

A Process for Using Big Data for Security: Interpret and Identify



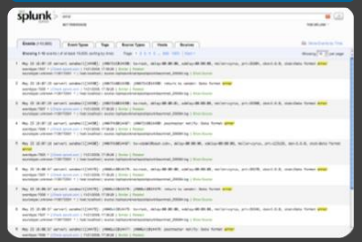
- What are the mitigating factors?
- Does the end of the quarter cause increased access to financial data?
- Does our statistical model need to change due to network architecture changes, employee growth, etc?
- Can we gather vacation information to know when it is appropriate for HPA users to access data from foreign soil.
- What are the changes in attack patterns?

Big Data Platform: Insight for Business Risk

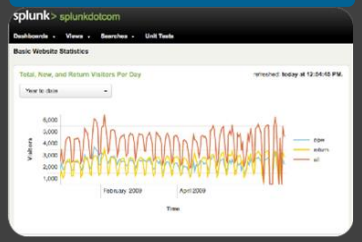
Business Risk and Security

Security & Compliance

IT Operations Management



Business Analytics



Web Intelligence

Application Monitoring


Thank You
splunk>

Outside Live Threat Intelligence

- Live data sampling from 38 international data centers
- Presence in top 20 Internet Exchange (IX) points world wide
- Core Long haul fiber access from tier 1 operators with several 10 Gbps pipes
- 1500 factors for creating an IPQ risk score to asses potential attacks